# Eliminating spin contamination in auxiliary-field quantum Monte Carlo: realistic potential energy curve of $F_2$

Wirawan Purwanto, W. A. Al-Saidi,* Henry Krakauer, and Shiwei Zhang
*Department of Physics, College of William and Mary, Williamsburg, Virginia 23187-8795, USA*
(Dated: February 2, 2008)

The use of an approximate reference state wave function $|\Phi_r\rangle$ in electronic many-body methods can break the spin symmetry of Born-Oppenheimer spin-independent Hamiltonians. This can result in significant errors, especially when bonds are stretched or broken. A simple spin-projection method is introduced for auxiliary-field quantum Monte Carlo (AFQMC) calculations, which yields spin-contamination-free results, even with a spin-contaminated $|\Phi_r\rangle$. The method is applied to the difficult $F_2$ molecule, which is unbound within unrestricted Hartree-Fock (UHF). With a UHF $|\Phi_r\rangle$, spin contamination causes large systematic errors and long equilibration times in AFQMC in the intermediate, bond-breaking region. The spin-projection method eliminates these problems, and delivers an accurate potential energy curve from equilibrium to the dissociation limit using the UHF $|\Phi_r\rangle$. Realistic potential energy curves are obtained with a cc-pVQZ basis. The calculated spectroscopic constants are in excellent agreement with experiment.

PACS numbers: 71.15.-m,02.70.Ss,31.25.-v,31.15.Ar

## I. INTRODUCTION

A standard approach in many-body electronic structure methods is to obtain ground and excited state energies from an approximate reference state wave function $|\Phi_r\rangle$. For example, the coupled-cluster (CC) approximation with single, double, and perturbative triple excitations [CCSD(T)], which is widely available in quantum chemistry computer codes, typically uses a Hartree-Fock (HF) single-determinant $|\Phi_r\rangle$.[1] Ground state quantum Monte Carlo (QMC) stochastic methods,[2,3,4,5] which are exact in principle, use projection from any $|\Phi_r\rangle$ that has non-zero overlap with the ground state wave function (WF). In practice, however, the Fermionic sign problem[4,5,6,7,8] must be controlled to achieve accurate results. Diffusion QMC (DMC) uses a single- or multi-reference WF to impose approximate Fermionic nodal boundary conditions in real space and also includes a Jastrow factor to reduce the stochastic variance. The recently developed phaseless auxiliary-field quantum Monte Carlo (AFQMC) method[5,9,10,11] is an alternative and complementary QMC approach, which samples the many-body wave function with random walkers in the space of Slater determinants. AFQMC provides a different route to controlling the sign problem, using the complex overlap of the walkers with $|\Phi_r\rangle$, which is frequently just a single HF determinant. Like the CC method, the AFQMC method works in a chosen single-particle basis, and it has been successfully applied using Gaussian[10,12,13,14] and plane wave[5,9,11] basis sets.

While these correlated methods are generally quite accurate near equilibrium geometries, the use of an approximate $|\Phi_r\rangle$ can introduce uncontrolled errors as bonds are stretched or broken.[15,16,17,18] The main reason for this is that correlation effects become increasingly important in the transition region where a system begins to dissociate into its fragments, which are themselves often open shell systems. The quality of $|\Phi_r\rangle$ typically degrades in this region, since it is derived from a simple level of theory. A second reason is the breaking of spin or spatial symmetries in these simple reference WFs.

In previous applications, phaseless AFQMC with an unrestricted Hartree-Fock (UHF) single-determinant $|\Phi_r^{\rm UHF}\rangle$ was found to often give better overall and more uniform accuracy than CCSD(T) in mapping the potential-energy curve (PEC).[10,12,13] In some cases, however, such as the BH and $N_2$ molecules, achieving quantitative accuracy of a few m$E_{\rm h}$ for the entire PEC required multi-determinant $|\Phi_r\rangle$.[14] In these cases, spin contamination did not appear to be a major source of the error seen in the calculations with UHF reference states. In this paper, we show that, with a single-determinant $|\Phi_r^{\rm UHF}\rangle$, the AFQMC potential energy curve of the difficult $F_2$ molecule is qualitatively incorrect in the intermediate dissociation region. Spin contamination of $|\Phi_r^{\rm UHF}\rangle$ is found to be the dominant factor for this error. We describe a simple spin-projection method to effectively remove spin-contamination effects.

With $|\Phi_r^{\rm UHF}\rangle$ and the spin-projection method, the AFQMC results of $F_2$ are shown to be accurate (within a few m$E_{\rm h}$ of the near-exact CCSDTQ result) across the entire PEC. One of the main appeals of QMC methods is that the computational cost typically scales with systems size as a low power. Using larger basis sets (cc-pVTZ and cc-pVQZ), we then obtain realistic PECs and spectroscopic constants and compare them with experimental results.

The remainder of the paper is organized as follows. Section II discusses the difficulties in calculating accurate $F_2$ PECs. In Section III a simple method is described that removes spin-contamination effects in AFQMC calculations. Realistic $F_2$ potential energy curves and spectroscopic constants are presented in Section IV. Finally, Section V summarizes and discusses our principal results.

## II. SPIN CONTAMINATION EFFECTS IN THE DISSOCIATION OF THE $F_2$ MOLECULE

The difficulty in treating the dissociation of the $F_2$ molecule is already evident at the mean-field level of theory. The upper panel of Fig. 1 shows PECs from HF and density functional theory (DFT). UHF does not predict a bound molecule,[21,22]

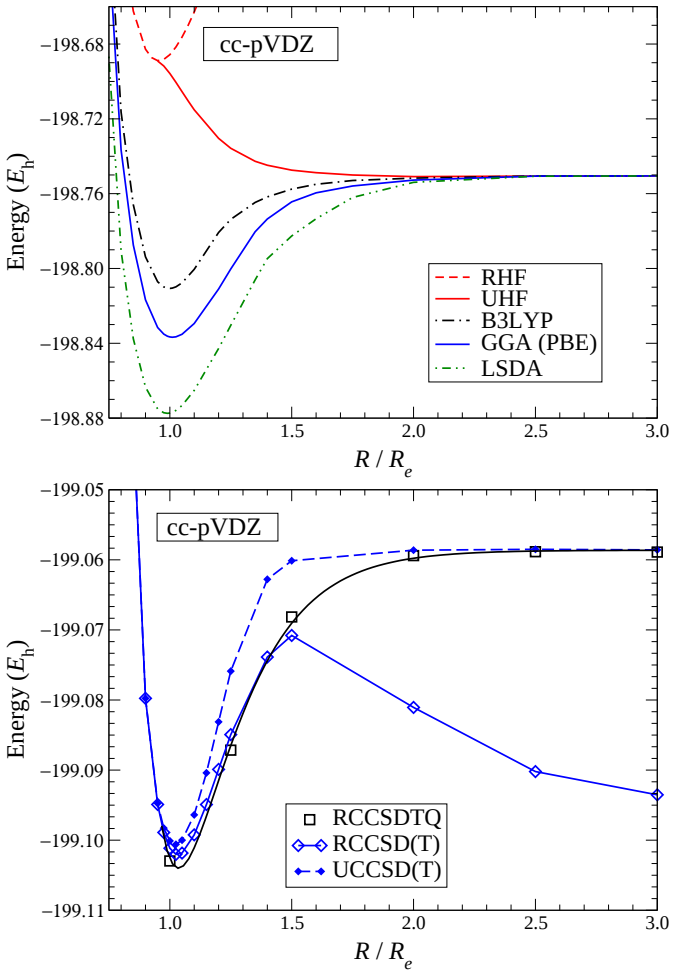


FIG. 1: (Color online) PECs for $F_2$ using the cc-pVDZ basis. The F–F internuclear distance, $R$, is shown in units of the equilibrium value $R_e \equiv 1.41193\,\text{Å}$.[19] (Note the energy scales are different between the upper and lower figures.) Upper figure: mean-field results, with constant shifts added to the DFT energies so that all energies match the UHF energy at $R/R_e = 3.0$. Lower figure: coupled-cluster results, both spin-restricted [RCCSD(T) and RCCSDTQ] and spin-unrestricted [UCCSD(T)]. The RCCSDTQ results are from Ref. 20. A Morse fit through the RCCSDTQ points is shown as a guide to the eye. Straight line segments connect the RCCSD(T) and UCCSD(T) points.

while the restricted HF (RHF) curve is artificially bound with a minimum that is $5\%$ too low compared to experiment. The DFT local spin-density approximation[23] (LSDA) and generalized gradient approximation[24] (GGA/PBE) yield dissociation energies which are too large. The hybrid B3LYP[25,26] dissociation energy is closer to experiment, but the shape of the B3LYP PEC is not correct in the intermediate region (see below in Sec. IV). All our HF and DFT calculations were carried out using the quantum chemistry computer program GAUSSIAN98.[27]

The difficulty of treating $F_2$ dissociation, even using correlated methods, is illustrated in the bottom panel of Fig. 1. For the small (cc-pVDZ) basis set[28,29] chosen here, the spin-restricted coupled-cluster including up to quadrupole excitations (RCCSDTQ) is within reach, which is expected to give close to exact results in this case. The RCCSD(T) and UCCSD(T) calculations, done with GAUSSIAN98[27] or NWCHEM,[30] use a single-determinant RHF or UHF $|\Phi_r\rangle$, respectively. The RCCSD(T) method breaks down in the dissociation limit, where the RHF $|\Phi_r\rangle$ is very poor (as seen in the upper panel of the figure). The UCCSD(T) PEC is accurate in the dissociation limit, but its shape begins to be distorted near the equilibrium bond length and shows significant error in the intermediate region.

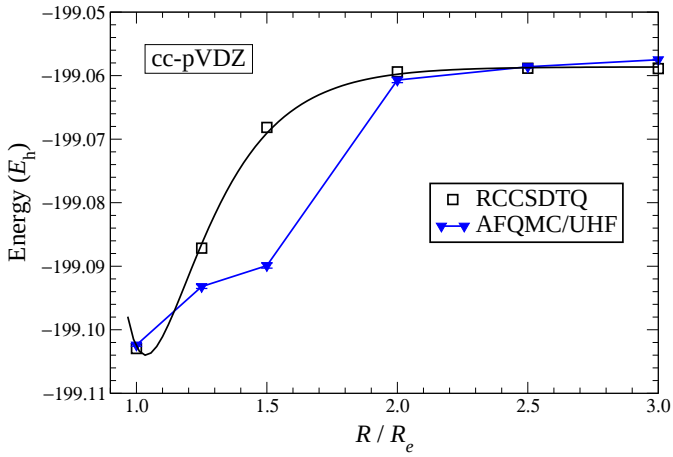


FIG. 2: (Color online) AFQMC $F_2$ PEC using a UHF reference state WF, compared to RCCSDTQ results from Fig. 1. The cc-pVDZ basis set is used. A Morse fit through the RCCSDTQ points is shown as a guide to the eye. Straight line segments connect the AFQMC/UHF points. The large deviations at $R/R_e \sim 1.5$ are due to spin contamination of the UHF reference state WF (see text).

The AFQMC PEC calculated with $|\Phi_r^{\text{UHF}}\rangle$ (labeled AFQMC/UHF) shows good agreement with RCCSDTQ near equilibrium and in the dissociation limit, as seen in Fig. 2. In the intermediate regime, however, AFQMC/UHF shows deviations of more than $20\,\text{m}E_\text{h}$. The poor results in this regime are due to the AFQMC phase-free approximation[5] when it is applied to a walker population that is spin-contaminated. The approximation, which depends on the accuracy of $|\Phi_r\rangle$, is analogous to that in the fixed node DMC method, whose performance depends on the accuracy of the $|\Phi_r\rangle$ nodal hypersurface. In view of the inability of $|\Phi_r^{\text{UHF}}\rangle$ to even bind $F_2$ and its poor quality in the intermediate regime, the inaccurate AFQMC results are perhaps not too surprising.

A brute force approach to improve the AFQMC PEC is to use a better $|\Phi_r\rangle$, through the use of a multi-determinant reference wave function. Indeed, using a generalized valence bond (GVB)[31] or complete active space self consistent field (CASSCF)[32] $|\Phi_r\rangle$ in AFQMC (labeled AFQMC/GVB and AFQMC/CASSCF, respectively) eliminates most of the error, as shown in Fig. 3. The GVB WF is a perfect-pairing GVB(1/2) wave function, where the electron pair responsible for the chemical bonding in $F_2$ (those in the $2p_z\,\sigma_g$ orbital in RHF) now occupy a pair of nonorthogonal, $2p_z$-atomic-like orbitals. The GVB WF has the proper dissociation limit.

The CASSCF(10,12) is for 10 active electrons and an active space of 12 molecular orbitals. The CASSCF WF is truncated, retaining only those determinants whose weights (the square of the configuration-interaction coefficients) are greater than $4 \times 10^{-4}$. (The adequacy of this cutoff was tested by performing additional calculations including determinants with weights larger than $10^{-4}$ in the trial WF. Within statistical errors, QMC energies similar to those with the higher weight cutoff were obtained.) The computational cost in AFQMC with a multi-determinant $|\Phi_r\rangle$ scales linearly with the number of determinants, although the real cost is typically less since a better $|\Phi_r\rangle$ reduces statistical errors.[14] In the next section, we show that the improved AFQMC/GVB and AFQMC/CASSCF PECs are largely due to the elimination of errors from spin contamination.

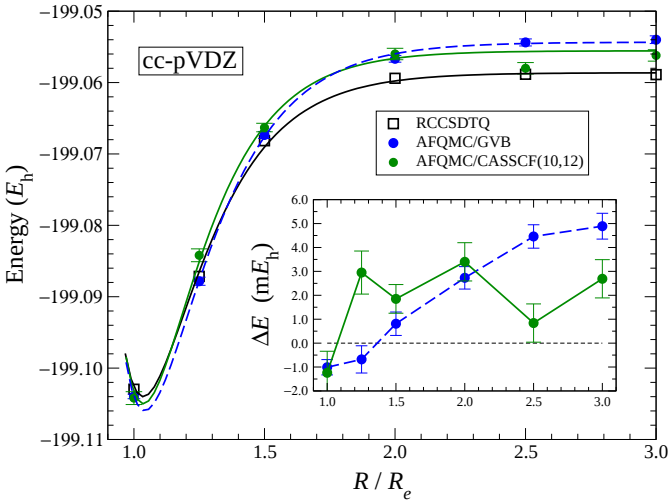


FIG. 3: (Color online) Improvement of the AFQMC $F_2$ PEC using multi-determinant reference state WFs from GVB(1/2) and CASSCF(10,12). RCCSDTQ from Fig. 1 is shown for comparison. Morse fits are shown as a guide to the eye. The inset shows energy differences $\Delta E$ (in m$E_h$) compared to RCCSDTQ. All calculations use the cc-pVDZ basis set.

## III. ELIMINATING SPIN CONTAMINATION IN AFQMC: SPIN PROJECTION METHOD

While the exact eigenstates of a spin-independent nonrelativistic electronic Hamiltonian are eigenstates of the total spin operator $\hat{S}^2$ and its $z$ component $\hat{S}_z$, approximate wave functions may not be eigenstates of $\hat{S}^2$ unless special care is taken. Such approximate wave functions are called *spin contaminated*. For simplicity, we restrict the discussion in this section to the case where the reference state wave function is given by $|\Phi_r^{\rm UHF}\rangle$ with $S_z = 0$, approximating the exact singlet ground state $|\Phi_0^s\rangle$. In this case $|\Phi_r^{\rm UHF}\rangle$ will generally be spin contaminated, *i.e.*, containing triplet $|\Psi^t\rangle$ and higher spin states:

$$|\Phi_r^{\rm UHF}\rangle = c_s|\Psi^s\rangle + c_t|\Psi^t\rangle + \dots , \quad (1)$$

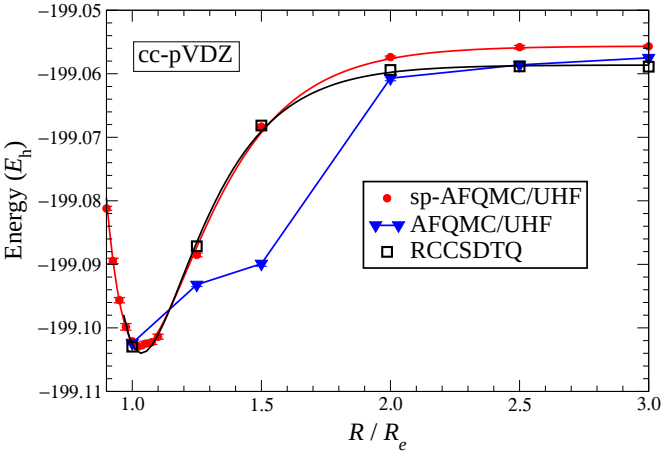


FIG. 4: (Color online) Improvement of the AFQMC $F_2$ PEC using spin projection with a single determinant UHF reference state wave function. Spin-projected (sp-AFQMC/UHF) results are compared to standard AFQMC/UHF without spin projection and to RCCSDTQ results from Fig. 2. Morse fits are shown as a guide to the eye, except for AFQMC/UHF. All calculations use the cc-pVDZ basis.

where $|\Psi^s\rangle$ is a linear combination of the ground and excited singlet states. In the UHF result of Fig. 1, the expectation value of the total electronic spin operator $\hat{S}^2$ in $|\Phi_r^{\rm UHF}\rangle$ is 0.362, 0.978, and 1.004 at $R/R_e = 1.0$, 1.5, and 3.0, respectively, indicating a high level of spin contamination in which the triplet component grows as the molecule is stretched.

Ideally, AFQMC projection of $|\Phi_r^{\rm UHF}\rangle$ would lead to the exact spin-contamination-free ground state, since

$$\left(e^{-\tau\hat{H}}\right)^n |\Phi_r^{\rm UHF}\rangle \rightarrow C_0|\Phi_0^s\rangle + C_1 e^{-n\tau(E_1-E_0^s)}|\Phi_1\rangle , \quad (2)$$

where $|\Phi_1\rangle$ is the exact first excited state, $\tau$ is the time-step parameter, and as $n \rightarrow \infty$, all components except $|\Phi_0^s\rangle$ become vanishingly small. The use of the phase-free approximation,[5] however, effectively modifies this projection so that a triplet component can survive. Thus, the population of AFQMC random walkers will be spin contaminated if it was initialized with $|\Phi_r^{\rm UHF}\rangle$. In $F_2$, the presence of a nearby triplet state[33] at bond lengths $R/R_e \gtrsim 1.4$ exacerbates this, and this is where the AFQMC/UHF PEC shows the largest error.

In the previous section, spin contamination in AFQMC was eliminated through the use of a (nearly) spin-pure multi-determinant $|\Phi_r\rangle$, which effectively filters the population of random walkers, retaining only the spin-pure component regardless of how the population was initialized. The GVB $|\Phi_r\rangle$ is spin-contamination free by design, and the truncated CASSCF $|\Phi_r\rangle$ is nearly free of spin contamination. The elimination/reduction of spin contamination in the GVB and truncated CASSCF $|\Phi_r\rangle$ is a main factor in the improvement of the corresponding QMC results. A strong clue to this is seen in the case with GVB $|\Phi_r\rangle$, where the GVB WF has only two determinants and has a variational energy within $\sim 1$ m$E_h$ that of the UHF at $R/R_e = 1.5$ (see Table I), and yet QMC/GVB greatly improves over QMC/UHF.

A simpler way to eliminate spin contamination in AFQMC

TABLE I: Comparison of computed $F_2$ PEC for various methods, using the cc-pVDZ basis. The RCCSDT and RCCSDTQ results are from Ref. 20. Energies are in $E_h$. QMC statistical errors are on the last digit and are shown in parentheses.

| | $R/R_e$ | | | | | |
|---|---|---|---|---|---|---|
| | 1.0 | 1.25 | 1.5 | 2.0 | 2.5 | 3.0 |
| RHF | −198.685670 | −198.612171 | −198.527711 | −198.419839 | −198.374025 | −198.355748 |
| UHF | −198.695746 | −198.735754 | −198.747441 | −198.750892 | −198.750597 | −198.750518 |
| GVB | −198.761466 | −198.759320 | −198.748618 | −198.743801 | −198.743570 | −198.743650 |
| CASSCF(10,12) | −198.886738 | −198.874892 | −198.857896 | −198.850284 | −198.849831 | −198.849732 |
| RCCSD(T) | −199.101152 | −199.084940 | −199.070790 | −199.081058 | −199.090213 | −199.093534 |
| UCCSD(T) | −199.100100 | −199.075878 | −199.060126 | −199.059302 | −199.058784 | −199.058687 |
| RCCSDT | −199.101417 | −199.084493 | −199.065170 | −199.057558 | −199.058023 | −199.057933 |
| RCCSDTQ | −199.102961 | −199.087149 | −199.068153 | −199.0594 | −199.05884 | −199.05889 |
| AFQMC/UHF | −199.1024(2) | −199.0932(3) | −199.0899(4) | −199.0607(4) | −199.0586(2) | −199.0575(1) |
| AFQMC/GVB | −199.1040(3) | −199.0878(6) | −199.0673(5) | −199.0567(5) | −199.0544(5) | −199.0540(5) |
| AFQMC/CASSCF(10,12) | −199.1042(9) | −199.0842(9) | −199.0663(6) | −199.0560(8) | −199.0580(8) | −199.0562(8) |
| sp-AFQMC/UHF | −199.1020(5) | −199.0876(7) | −199.0686(7) | −199.0574(2) | −199.0558(3) | −199.0562(5) |

is to ensure that the population of random walkers consists of spin-pure (*i.e.* RHF-type) Slater determinants, $\{\,|\phi_s\rangle\,\}$. Almost all phaseless AFQMC electronic structure calculations to date[5,10,11,13] have used Hubbard-Stratonovich (HS) transformations which preserve spin symmetry,[34]

$$\hat{v}_{\rm HS}(\mathbf{x}) = \hat{v}_\uparrow(\mathbf{x}) + \hat{v}_\downarrow(\mathbf{x})\,, \tag{3}$$

where $\mathbf{x}$ denotes HS auxiliary fields, and the one-body operators $\hat{v}_\uparrow(\mathbf{x})$ and $\hat{v}_\downarrow(\mathbf{x})$ have identical forms. For example, in the plane-wave formalism,[5,11] $\hat{v}_{\rm HS}(\mathbf{x})$ is essentially a Fourier component of the density operator. Thus, if a random walker is initialized to a spin-pure Slater determinant with $S_z = 0$, its spin state cannot be modified by the QMC propagation $|\phi_s'\rangle = e^{-\hat{v}_{\rm HS}(\mathbf{x})}|\phi_s\rangle$. Typically, the same trial WF is used in phaseless AFQMC to generate the initial population, to guide the importance sampling, and to impose the phaseless constraint.[5,35] This of course does not have to be the case. Here we use a spin-pure state to initialize the walkers. Since each walker in the population $\{\,|\phi_s\rangle\,\}$ remains spin-pure, the local energy $E_L[\phi_s]$ projects out the triplet and higher components of $|\Phi_r^{\rm UHF}\rangle$:

$$E_L[\phi_s] = \frac{\langle\Phi_r^{\rm UHF}|\hat{H}|\phi_s\rangle}{\langle\Phi_r^{\rm UHF}|\phi_s\rangle}\,. \tag{4}$$

The mixed estimator for the ground state energy is determined by the local energy, so it too is spin-uncontaminated. Thus, higher spin states have no effect on either the AFQMC projection, the phase-free approximation, or the ground state energy estimation.

The spin-projected AFQMC (sp-AFQMC) method described above shows a dramatic improvement over the spin-contaminated AFQMC/UHF in $F_2$, as seen in Fig. 4. In the sp-AFQMC/UHF calculations, the walker population is initialized with the RHF solution $|\phi_s\rangle = |{\rm RHF}\rangle$, but $|\Phi_r^{\rm UHF}\rangle$ is used to implement the phase-free constraint[5] and to calculate the local energy. Table I tabulates the energies for all methods using the cc-pVDZ basis. We see that the sp-AFQMC/UHF PEC is in excellent agreement with the RCCSDTQ result, with a maximum discrepancy of $\sim 3\,{\rm m}E_h$. This accuracy is in fact slightly better than that of either AFQMC/CASSCF or AFQMC/GVB.

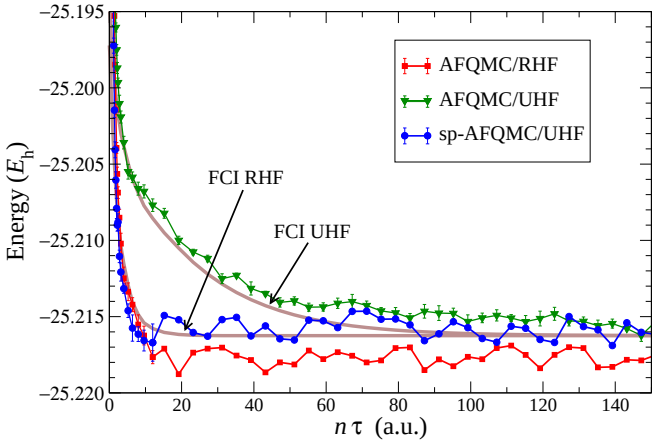


FIG. 5: (Color online) Slow energy equilibration as a function of imaginary time $n\tau$ of spin contaminated AFQMC/UHF, compared to spin contamination free AFQMC/RHF and sp-AFQMC/UHF for the BH molecule at $R_e = 1.2344$ Å. The cc-pVDZ basis is used. To reduce clutter, QMC statistical errors are not shown in sp-AFQMC/UHF and AFQMC/RHF for $n\tau > 15$, but the average size of the error bar is indicated for each in the legend. FCI-derived RHF and UHF projection curves (see text) are also shown, calculated using Eq. (2). The FCI ground-state energy is $-21.216\,249\,E_h$.

In addition to removing spurious spin contamination effects in the calculated AFQMC energy, sp-AFQMC can sometimes also reduce the imaginary time [see Eq. (2)] needed to obtain energy equilibration. This is illustrated in Fig. 5 for the BH molecule. The spin-contaminated AFQMC/UHF has an equilibration time $n\tau \sim 100$ a.u., about an order of magnitude larger than the spin-contamination-free sp-AFQMC/UHF ($n\tau \sim 10$ a.u.). For comparison, the curve from AFQMC using $|\Phi_r^{\rm RHF}\rangle$ is also shown. Starting from the same initial state, the the spin-contamination-free AFQMC/RHF has a short equilibration time similar to sp-AFQMC/UHF, but the converged result has a larger systematic error, because of the poorer quality of $|\Phi_r^{\rm RHF}\rangle$ as the constraining WF in the phase-

less approximation. The different behaviors of the equilibration time can be understood by comparing with FCI-derived RHF and UHF projections, which are shown in Fig. 5. We calculate the "exact" projection results by expanding the UHF and RHF initial WFs in terms of a truncated set of the FCI eigenstates (the first 80 eigenstates, obtained with GAMESS[36]). With a UHF initial WF, the long equilibration time is due to the presence of low-lying triplet components [see Eq. (1)], which results in smaller effective gap $(E_1 - E_0^s)$ in Eq. (2). The RHF WF, on the other hand, has no overlap with any triplet state, and consequently the effective gap is larger and the equilibration time shorter.

TABLE II: The PEC of $F_2$, computed using spin-projected phaseless AFQMC with UHF trial wave function in cc-pVTZ and cc-pVQZ basis. Energies are in $E_h$. QMC statistical errors are on the last digit and are shown in parentheses.

| $R/R_e$ | cc-pVTZ | cc-pVQZ |
|---|---|---|
| 0.80 | −199.2321(4) | −199.3372(5) |
| 0.85 | −199.2805(5) | −199.3824(6) |
| 0.9 | −199.3061(6) | −199.4099(3) |
| 0.925 | −199.3134(4) | −199.4159(6) |
| 0.95 | −199.3186(5) | −199.4218(7) |
| 0.975 | −199.3210(6) | −199.4231(5) |
| 1.0 | −199.3211(4) | −199.4241(4) |
| 1.025 | −199.3209(4) | −199.4236(4) |
| 1.05 | −199.3190(4) | −199.4219(4) |
| 1.075 | −199.3176(5) | −199.4201(4) |
| 1.1 | −199.3136(5) | −199.4170(4) |
| 1.25 | −199.2962(6) | −199.3975(4) |
| 1.4 | −199.2797(6) | −199.3807(3) |
| 1.5 | −199.2731(6) | −199.3732(4) |
| 1.75 | −199.2632(4) | −199.3641(2) |
| 2.0 | −199.2603(6) | −199.3605(4) |
| 3.0 | −199.2590(6) | −199.3593(4) |

## IV. REALISTIC $F_2$ POTENTIAL ENERGY CURVE: BASIS-SET CONVERGED SPIN-PROJECTED AFQMC RESULTS

We have shown that the sp-AFQMC PEC is accurate at the double zeta cc-pVDZ level, where near-exact CCSDTQ coupled-cluster results are available for comparison. As a function of bond stretching, sp-AFQMC delivers more uniform accuracy than RCCSD(T) and UCCSD(T) for the difficult $F_2$ molecule, with absolute errors of a few m$E_h$ or less. In this section, we employ large basis sets to obtain a realistic PEC. We also compute $F_2$ spectroscopic constants and compare them with experimental results.

Figure 6 presents the PECs of $F_2$ computed using sp-AFQMC/UHF for cc-pVTZ and cc-pVQZ basis sets.[29] For comparison, PECs from B3LYP, RCCSD(T) and UCCSD(T) are also shown, representing the best current theoretical results. (The B3LYP curves were shifted to agree with sp-AFQMC/UHF at $R/R_e = 3$.) The sp-AFQMC/UHF energies corresponding to Fig. 6 are also tabulated in Table II.

Computed spectroscopic constants are given in Table III together with those from the many-body RCCSD(T) and UCCSD(T), and the independent-electron LSDA, GGA/PBE, and B3LYP methods. The spectroscopic constants were obtained by fitting the calculated PECs in the range $0.8 \leq R/R_e \leq 1.25$ to a three-term extended Morse curve[39]

$$E(r) = E_0 + \sum_{n=2}^{4} \frac{C_n}{a^n} \left[1 - e^{-a(r-r_e)}\right]^n . \qquad (5)$$

The fitting procedure yields the molecular electronic energy, $E_0 \equiv E(r_e)$, equilibrium bond length $r_e$, and the harmonic frequency $\omega_0 = \sqrt{C_2/2\mu}$, where $\mu$ is the reduced mass of the $F_2$ molecule. The dissociation energy is given by $D_e \equiv E(3R_e) - E(r_e)$. For comparison, $D_e$ calculated from $2E(\text{atom}) - E(r_e)$, where $E(\text{atom})$ is a well-converged energy for the isolated atom, is also shown for the many-body results in the TZ and QZ basis sets.

The values of sp-AFQMC/UHF $r_e$ and $\omega_0$ in Table III are in excellent agreement with experiment. The dissociation energy $D_e = E(3R_e) - E(r_e)$ is overestimated, however. This is due to the overestimation of the total energy at large $R/R_e = 3$, which reflects the deficiency of a simple UHF $|\Phi_r\rangle$ in AFQMC for open-shell systems, as previously noted.[11] To obtain a more accurate $D_e$, an AFQMC calculation was performed for the isolated F atom with a truncated CASSCF(7,13) $|\Phi_r\rangle$. The $2s$ through $3d$ orbitals were included in the active space of the CASSCF WF. The truncation retains determinants with weight greater than $2 \times 10^{-4}$, resulting in a $|\Phi_r\rangle$ with 47 determinants. In cc-pVQZ, the atomic energy thus calculated is $E(\text{atom}) = -99.6811(5)\, E_h$, while the corresponding RCCSD(T) and UCCSD(T) values are $-99.681\,704$ and $-99.681\,576\, E_h$, respectively. The dissociation energy obtained with $D_e = 2E(\text{atom}) - E(r_e)$ is in excellent agreement with experiment.

The variations in the results from the TZ to the QZ basis sets are still visible but quite small (especially in $r_e$ and $\omega_0$). It is thus reasonable to expect the residual finite basis set error to be small in the QZ basis. A simple extrapolation to the infinite basis limit[40] increases $D_e$ only by $0.02$ eV ($0.7\,\text{m}E_h$) from the cc-pVQZ value. The shape of the sp-AFQMC PEC should thus be very close to that at basis set convergence. (In contrast, the residual error of the cc-pVQZ absolute molecular energies is approximately $110\,\text{m}E_h$, estimated using nonrelativistic energies published in the literature.[37,41])

Compared to the sp-AFQMC/UHF PEC, the RCCSD(T) and UCCSD(T) PECs in Fig. 6 show the same shortcomings as seen with the cc-pVDZ basis in Fig. 1. While the RCCSD(T) PEC near equilibrium is in good agreement with sp-AFQMC/UHF, it is very poor in the dissociation limit. For this reason, the RCCSD(T) dissociation energy $D_e$ shown in Table III is computed only from $D_e = 2E(\text{atom}) - E(r_e)$. The UCCSD(T) PEC is accurate in the dissociation limit, but its shape is significantly distorted near equilibrium. Consequently, the UCCSD(T) spectroscopic constants are not in as good agreement with experiment. The RCCSD(T) $r_e$ and $\omega_0$ are also in excellent agreement with experiment, while the UCCSD(T) $\omega_0$ is $13\%$ too large. This is consistent with

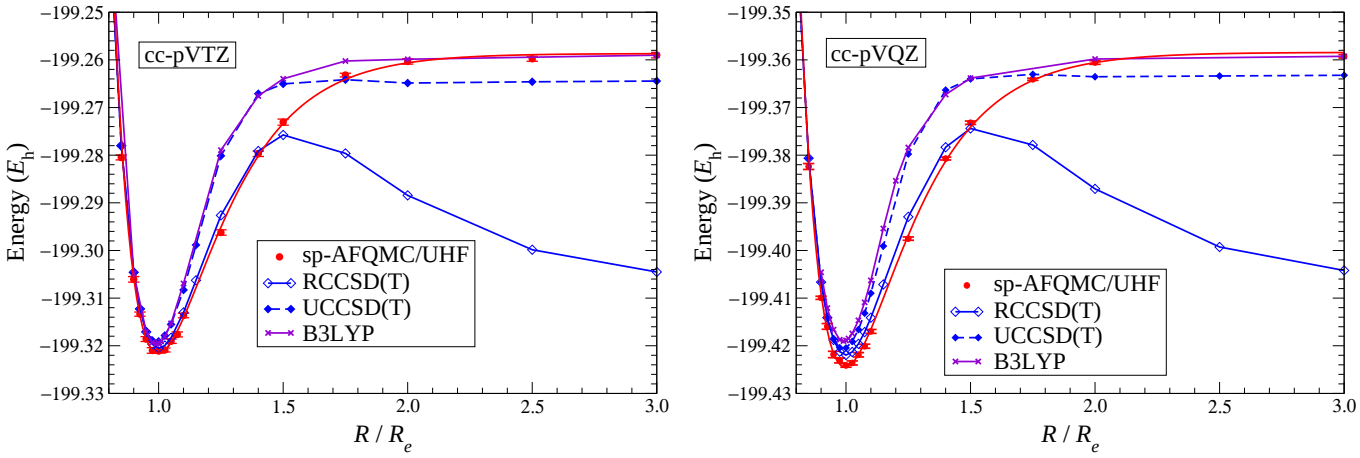


FIG. 6: (Color online) PECs for $F_2$ with cc-pVTZ and cc-pVQZ basis sets. A Morse fit passes through the sp-AFQMC/UHF points. Straight line segments connect results of the other methods. The B3LYP curves were shifted to agree with sp-AFQMC/UHF at $R/R_e = 3$.

TABLE III: Computed $F_2$ spectroscopic constants for three basis sets, together with experimental results.

| | Expt[a] | AFQMC | RCCSD(T) | UCCSD(T) | LSDA | GGA/PBE | B3LYP |
|---|---|---|---|---|---|---|---|
| Basis: cc-pVDZ | | | | | | | |
| $r_e$ (Å) | 1.4131(8) | 1.467(4) | 1.4571 | 1.4428 | 1.3970 | 1.4276 | 1.4097 |
| $\omega_0$ ($cm^{-1}$) | 916.64 | 725(36) | 785 | 853 | 1026 | 958 | 1033 |
| $D_e$ (eV)[b] | 1.693(5) | 1.293(7) | 1.180[c] | 1.142 | 3.454 | 2.347 | 1.638 |
| Basis: cc-pVTZ | | | | | | | |
| $r_e$ (Å) | 1.4131(8) | 1.411(3) | 1.4131 | 1.3987 | 1.3863 | 1.4138 | 1.3957 |
| $\omega_0$ ($cm^{-1}$) | 916.64 | 928(30) | 926 | 1022 | 1065 | 1001 | 1072 |
| $D_e$ (eV)[b] | 1.693(5) | 1.70(2) | – | 1.493 | 3.486 | 2.351 | 1.651 |
| $D_e$ (eV)[c] | 1.693(5) | 1.60(1) | 1.523 | 1.495 | | | |
| Basis: cc-pVQZ | | | | | | | |
| $r_e$ (Å) | 1.4131(8) | 1.411(2) | 1.4108 | 1.3946 | 1.3856 | 1.4136 | 1.3944 |
| $\omega_0$ ($cm^{-1}$) | 916.64 | 912(11) | 929 | 1036 | 1062 | 997 | 1109 |
| $D_e$ (eV)[b] | 1.693(5) | 1.77(1) | – | 1.567 | 3.473 | 2.321 | 1.634 |
| $D_e$ (eV)[c] | 1.693(5) | 1.70(1) | 1.594 | 1.569 | | | |

[a]The dissociation energy $D_e$ is from Ref. 37 (the zero point and spin-orbit energies have been removed). The equilibrium internuclear distance $r_e$ is from Ref. 38, and the vibrational frequency $\omega_0$ is from Ref. 19.
[b]The dissociation energy calculated using $E(3R_e) - E(r_e)$.
[c]The dissociation energy calculated using $2E(\text{atom}) - E(r_e)$. In AFQMC, $E(\text{atom})$ is calculated with a truncated CASSCF(7,13) $|\Phi_r\rangle$.

Fig. 1, where the UCCSD(T) potential well is too narrow compared with the near-exact RCCSDTQ result. Curiously, the UCCSD(T) and B3LYP PECs show very similar deviations near equilibrium.

As expected, LSDA, GGA/PBE, and B3LYP show more rapid convergence with basis set size than the correlated methods. The B3LYP $D_e$ is good, but since the shape of its PEC is incorrect, $\omega_0$ is $\sim 20\%$ too large and the equilibrium bond length is too small. (The large discrepancy here underscores the difficult nature of $F_2$; in other molecules, B3LYP results are typically found to be in good agreement with experiment.[42]) Both LSDA and GGA/PBE have poor $\omega_0$ and $D_e$, while their equilibrium bond lengths $r_e$ are within $\sim 2\%$ of experiment.

## V. SUMMARY AND DISCUSSION

The accuracy of AFQMC depends on the reference wave function $|\Phi_r\rangle$, which is used to implement the phase-free constraint.[5] This is analogous to DMC, which uses a reference $|\Phi_r\rangle$ to impose the fixed-node approximation to control the sign problem. In previous applications, AFQMC was found to have less reliance on the quality of $|\Phi_r\rangle$, and frequently a single-determinant $|\Phi_r\rangle$ was found adequate. In these cases, the best results were obtained using the best variational single determinant reference state, namely the HF solution when RHF and UHF are the same (e.g., in the $H_2O$ molecule at equilibrium[10]), or the UHF solution $|\Phi_r^{\text{UHF}}\rangle$ when the two differ. Moreover, the AFQMC method seemed relatively insensitive, within the spin unrestricted framework, to whether a HF,

DFT, or hybrid B3LYP Slater determinant was used as $|\Phi_r\rangle$.[12] In some cases, however, such as the BH and $N_2$ molecules, achieving quantitative accuracy of a few m$E_h$ for the entire PEC required multi-determinant $|\Phi_r\rangle$.[14]

It is shown here that, with $|\Phi_r^{\rm UHF}\rangle$, the AFQMC PEC of the difficult $F_2$ molecule is qualitatively incorrect in the intermediate dissociation region. Spin-contamination is identified as the primary source of the error. We have introduced a simple scheme, sp-AFQMC, that effectively removes spin-contamination effects, regardless of the choice of $|\Phi_r\rangle$. It is also illustrated how spin projection can often shorten the AFQMC equilibration time. $F_2$ calculations with sp-AFQMC/UHF were shown to give a PEC whose accuracy is better than a few m$E_h$ across the entire curve in the cc-pVDZ basis. To our knowledge, these are the most accurate results obtained by a theoretical method that easily scales up in system size. The full PEC curves from equilibrium to the dissociation limit were then calculated with cc-pVTZ and cc-pVQZ basis sets. Spectroscopic constants with the cc-pVQZ basis were found to be in excellent agreement with experiment.

The sp-AFQMC results with a single determinant $|\Phi_r^{\rm UHF}\rangle$ are comparable to those obtained with a multi-determinant $|\Phi_r\rangle$ trial WF from GVB or CASSCF. The spin-projection method thus further reduces the reliance of AFQMC on the choice of $|\Phi_r\rangle$, which is one of its most desirable features.

While the focus has been mainly on AFQMC using a single determinant $|\Phi_r^{\rm UHF}\rangle$ reference wave function, the method may also prove useful with multi-determinant $|\Phi_r\rangle$ with significant spin contamination. This could arise, for example, in treating correlated transition metal systems with truncated CASSCF wave functions.

### Acknowledgments

This work was supported by DOE/CMSN (DE-FG02-07ER46366), ONR (N000140110365 and N000140510055), NSF (DMR-0535529), and ARO (48752PH). Calculations were performed at the Center for Piezoelectrics by Design, and the College of William & Mary's SciClone cluster. We are grateful to Eric Walter for many useful discussions. The matrix elements used in our AFQMC calculations were obtained using a modified NWCHEM 4.6 code. The trial wave functions were obtained using NWCHEM and GAUSSIAN98 codes.

* Present address: Department of Physics, Cornell University, Ithaca, New York 14853, USA


1. R. J. Bartlett and M. Musiał, Reviews of Modern Physics **79**, 291 (pages 62) (2007).
2. D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
3. P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, J. Chem. Phys. **77**, 5593 (1982).
4. W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001), also see the references therein.
5. S. Zhang and H. Krakauer, Phys. Rev. Lett. **90**, 136401 (2003).
6. D. M. Ceperley and B. J. Alder, J. Chem. Phys. **81**, 5833 (1984).
7. S. Zhang and M. H. Kalos, Phys. Rev. Lett. **67**, 3074 (1991).
8. S. Zhang, in *Quantum Monte Carlo Methods in Physics and Chemistry*, edited by M. P. Nightingale and C. J. Umrigar (Kluwer Academic Publishers, 1999), cond-mat/9909090.
9. W. A. Al-Saidi, H. Krakauer, and S. Zhang, Phys. Rev. B **73**, 075103 (2006).
10. W. A. Al-Saidi, S. Zhang, and H. Krakauer, J. Chem. Phys. **124**, 224101 (2006).
11. M. Suewattana, W. Purwanto, S. Zhang, H. Krakauer, and E. J. Walter, Physical Review B (Condensed Matter and Materials Physics) **75**, 245123 (pages 12) (2007).
12. W. A. Al-Saidi, H. Krakauer, and S. Zhang, J. Chem. Phys. **125**, 154110 (2006).
13. W. A. Al-Saidi, H. Krakauer, and S. Zhang, J. Chem. Phys. **126**, 194105 (pages 8) (2007).
14. W. A. Al-Saidi, S. Zhang, and H. Krakauer, J. Chem. Phys. **127**, 144101 (2007).
15. E. R. Davidson and W. T. Borden, Journal of Physical Chemistry **87**, 4783 (1983).
16. C.-J. Huang, C. Filippi, and C. J. Umrigar, J. Chem. Phys. **108**, 8838 (1998).
17. J. S. Sears, C. D. Sherrill, and A. I. Krylov, J. Chem. Phys. **118**, 9084 (2003).
18. R. C. Lochan and M. Head-Gordon, J. Chem. Phys. **126**, 164101 (pages 11) (2007).
19. K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure. IV. Constants of Diatomic Molecules* (Van Nostrand Reinhold Company, 1979).
20. M. Musiał and R. J. Bartlett, J. Chem. Phys. **122**, 224102 (pages 9) (2005).
21. K. Hijikata, J. Chem. Phys. **34**, 221 (1961).
22. M. S. Gordon and D. G. Truhlar, Theor. Chim. Acta **71**, 1 (1987).
23. J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
24. J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
25. A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
26. P. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
27. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, et al., *Gaussian 98, Revision A.11.4*, Gaussian, Inc., Pittsburgh, PA (2002).
28. T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
29. K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. L. Windus, J. Chem. Inf. Model. **47**, 1045 (2007).
30. E. Aprà, T. Windus, T. Straatsma, E. Bylaska, W. de Jong, S. Hirata, M. Valiev, M. Hackler, L. Pollack, K. Kowalski, et al., *NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.6*, Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA (2004).
31. F. W. Bobrowicz and W. A. Goddard III, in *Methods of Electronic Structure Theory*, edited by H. F. Schaefer III (Plenum Press, New York, 1977), pp. 79–127.
32. M. W. Schmidt and M. S. Gordon, Annu. Rev. Phys. Chem. **49**, 233 (1998).
33. D. C. Cartwright and P. J. Hay, J. Chem. Phys. **70**, 3191 (1979).

[34] A spin (as opposed to charge) decomposition $(n_{i,\uparrow} - n_{i,\downarrow})$, which breaks spin symmetry, is found to be more efficient and is used in most lattice model constrained path Monte Carlo calculations (see, e.g., Ref. 35). This case will require further investigation.
[35] S. Zhang, J. Carlson, and J. E. Gubernatis, Phys. Rev. B **55**, 7464 (1997).
[36] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, et al., J. Comput. Chem. **14**, 1347 (1993).
[37] L. Bytautas and K. Ruedenberg, J. Chem. Phys. **122**, 154110 (2005).
[38] H. Edwards, E. Good, and D. Long, J. Chem. Soc. Faraday Trans. II **72**, 984 (1976).
[39] A. S. Coolidge, H. M. James, and E. L. Vernon, Phys. Rev. **54**, 726 (1938).
[40] D. Feller and K. A. Peterson, J. Chem. Phys. **108**, 154 (1998).
[41] C. Filippi and C. J. Umrigar, J. Chem. Phys. **105**, 213 (1996).
[42] M. O. Sinnokrot and C. D. Sherrill, J. Chem. Phys. **115**, 2439 (2001).